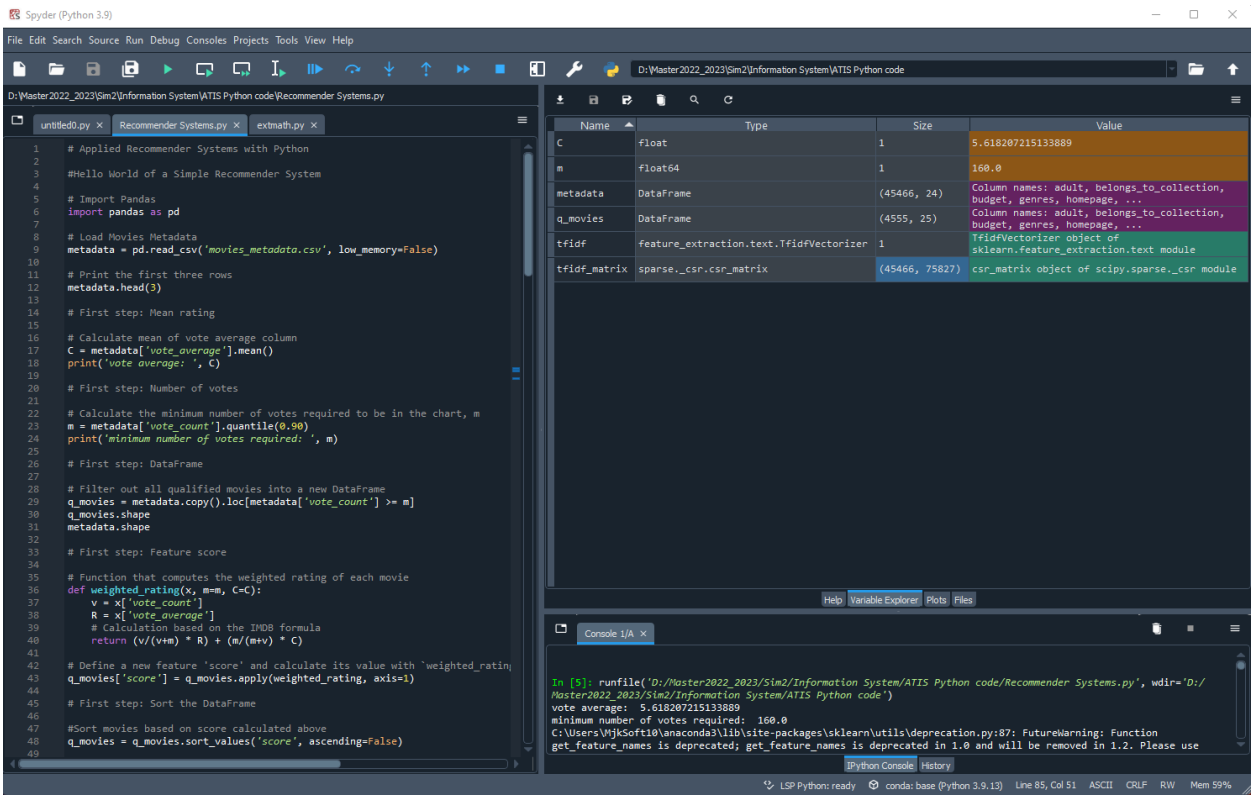
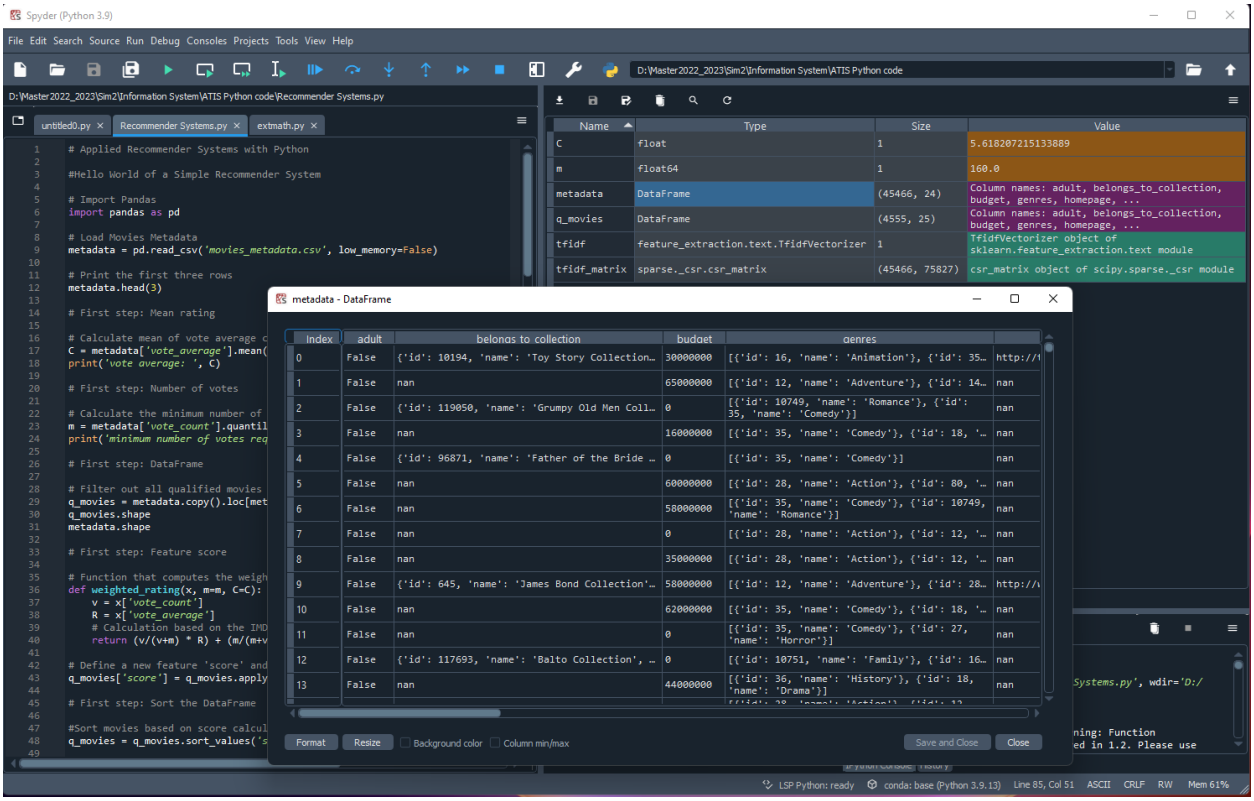
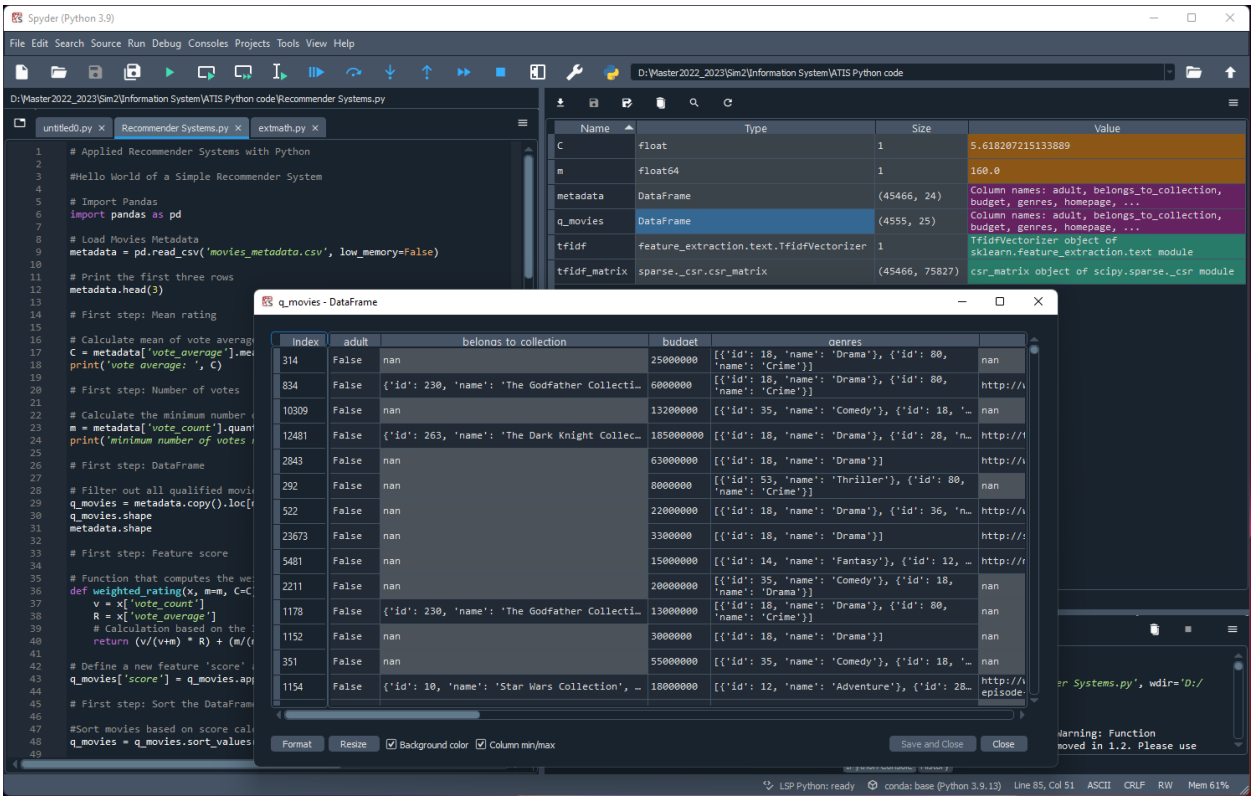

Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
D:\Master2022_2023\Sim2\Information System\ATIS Python code
D:\Master2022_2023\Sim2\Information System\ATIS Python code\Recommender Systems.py
untitled0.py
Recommender Systems.py
extmath.py
Applied Recommender Systems with Python
#Hello World of a Simple Recommender System
# Import Pandas
import pandas as pd
# Load Movies Metadata
metadata = pd.read_csv('movies_metadata.csv', low_memory=False)
# Print the first three rows
metadata.head(3)
# First step: Mean rating
# Calculate mean of vote average column
C = metadata['vote_average'].mean()
print('vote average: ', C)
# First step: Number of votes
# Calculate the minimum number of votes required to be in the chart, m
m = metadata['vote_count'].quantile(0.90)
print('minimum number of votes required: ', m)
# First step: DataFrame
# Filter out all qualified movies into a new DataFrame
q_movies = metadata.copy().loc[metadata['vote_count'] >= m]
q_movies.shape
metadata.shape
# First step: Feature score
# Function that computes the weighted rating of each movie
def weighted_rating(x, m=m, C=C):
v = x['vote_count']
R = x['vote_average']
# Calculation based on the IMDB formula
return (v/(v+m) * R) + (m/(m+v) * C)
Define a new feature 'score' and calculate its value with `weighted_rating
q_movies['score'] = q_movies.apply(weighted_rating, axis=1)
First step: Sort the DataFrame
#Sort movies based on score calculated above
q_movies = q_movies.sort_values('score', ascending=False)
Name Type Size Value
C float 1 5.618207215133889
m float64 1 160.0
metadata DataFrame (45466, 24) Column names: adult, belongs_to_collection, budget, genres, homepage, ...
q_movies DataFrame (4555, 25) Column names: adult, belongs_to_collection, budget, genres, homepage, ...
tfidf feature_extraction.text.TfidfVectorizer 1 TfidfVectorizer object of sklearn.feature_extraction.text module
tfidf_matrix sparse._csr.csr_matrix (45466, 75827) csr_matrix object of scipy.sparse._csr module
Help Variable Explorer Plots Files
Console 1/A
In [5]: runfile('D:/Master2022_2023/Sim2/Information System/ATIS Python code/Recommender Systems.py', wdir='D:/Master2022_2023/Sim2/Information System/ATIS Python code')
vote average: 5.618207215133889
minimum number of votes required: 160.0
C:\Users\MjkSoft10\anaconda3\lib\site-packages\sklearn\utils\deprecation.py:87: FutureWarning: Function get_feature_names is deprecated; get_feature_names is deprecated in 1.0 and will be removed in 1.2. Please use
IPython Console History
LSP Python: ready
conda: base (Python 3.9.13)
Line 85, Col 51
ASCII
CRLF
RW
Mem 59%


Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
D:\Master2022_2023\Sim2\Information System\ATIS Python code
D:\Master2022_2023\Sim2\Information System\ATIS Python code\Recommender Systems.py
metadata - DataFrame
Index adult belongs_to_collection budget genres
0 False {'id': 10194, 'name': 'Toy Story Collection… 30000000 [{'id': 16, 'name': 'Animation'}, {'id': 35… http://t
1 False nan 65000000 [{'id': 12, 'name': 'Adventure'}, {'id': 14… nan
2 False {'id': 119050, 'name': 'Grumpy Old Men Coll… 0 [{'id': 10749, 'name': 'Romance'}, {'id': 35, 'name': 'Comedy'}] nan
3 False nan 16000000 [{'id': 35, 'name': 'Comedy'}, {'id': 18, '… nan
4 False {'id': 96871, 'name': 'Father of the Bride … 0 [{'id': 35, 'name': 'Comedy'}] nan
5 False nan 60000000 [{'id': 28, 'name': 'Action'}, {'id': 80, '… nan
6 False nan 58000000 [{'id': 35, 'name': 'Comedy'}, {'id': 10749, 'name': 'Romance'}] nan
7 False nan 0 [{'id': 28, 'name': 'Action'}, {'id': 12, '… nan
8 False nan 35000000 [{'id': 28, 'name': 'Action'}, {'id': 12, '… nan
9 False {'id': 645, 'name': 'James Bond Collection'… 58000000 [{'id': 12, 'name': 'Adventure'}, {'id': 28… http://
10 False nan 62000000 [{'id': 35, 'name': 'Comedy'}, {'id': 18, '… nan
11 False nan 0 [{'id': 35, 'name': 'Comedy'}, {'id': 27, 'name': 'Horror'}] nan
12 False {'id': 117693, 'name': 'Balto Collection', … 0 [{'id': 10751, 'name': 'Family'}, {'id': 16… nan
13 False nan 44000000 [{'id': 36, 'name': 'History'}, {'id': 18, 'name': 'Drama'}] nan
Format Resize Background color Column min/max Save and Close Close
Name Type Size Value
C float 1 5.618207215133889
m float64 1 160.0
metadata DataFrame (45466, 24) Column names: adult, belongs_to_collection, budget, genres, homepage, ...
q_movies DataFrame (4555, 25) Column names: adult, belongs_to_collection, budget, genres, homepage, ...
tfidf feature_extraction.text.TfidfVectorizer 1 TfidfVectorizer object of sklearn.feature_extraction.text module
tfidf_matrix sparse._csr.csr_matrix (45466, 75827) csr_matrix object of scipy.sparse._csr module
LSP Python: ready
conda: base (Python 3.9.13)
Line 85, Col 51
ASCII
CRLF
RW
Mem 61%


Spyder (Python 3.9)
File Edit Search Source Run Debug Consoles Projects Tools View Help
D:\Master2022_2023\Sim2\Information System\ATIS Python code
D:\Master2022_2023\Sim2\Information System\ATIS Python code\Recommender Systems.py
untitled0.py
Recommender Systems.py
extmath.py
Applied Recommender Systems with Python
#Hello World of a Simple Recommender System
# Import Pandas
import pandas as pd
# Load Movies Metadata
metadata = pd.read_csv('movies_metadata.csv', low_memory=False)
# Print the first three rows
metadata.head(3)
# First step: Mean rating
# Calculate mean of vote averag
C = metadata['vote_average'].me
print('vote average: ', C)
# First step: Number of votes
# Calculate the minimum number
m = metadata['vote_count'].quant
print('minimum number of votes
# First step: DataFrame
# Filter out all qualified movi
q_movies = metadata.copy().loc[
q_movies.shape
metadata.shape
First step: Feature score
Function that computes the we
def weighted_rating(x, m=m, C=C
v = x['vote_count']
R = x['vote_average']
# Calculation based on the
return (v/(v+m) * R) + (m/(
Define a new feature 'score'
q_movies['score'] = q_movies.ap
First step: Sort the DataFram
#Sort movies based on score cal
q_movies = q_movies.sort_values
Name Type Size Value
C float 1 5.618207215133889
m float64 1 160.0
metadata DataFrame (45466, 24) Column names: adult, belongs_to_collection, budget, genres, homepage, ...
q_movies DataFrame (4555, 25) Column names: adult, belongs_to_collection, budget, genres, homepage, ...
tfidf feature_extraction.text.TfidfVectorizer 1 TfidfVectorizer object of sklearn.feature_extraction.text module
tfidf_matrix sparse._csr.csr_matrix (45466, 75827) csr_matrix object of scipy.sparse._csr module
q_movies - DataFrame
Index adult belongs_to_collection budget genres
314 False nan 25000000 [{'id': 18, 'name': 'Drama'}, {'id': 80, 'name': 'Crime'}] nan
834 False {'id': 230, 'name': 'The Godfather Collecti… 6000000 [{'id': 18, 'name': 'Drama'}, {'id': 80, 'name': 'Crime'}] http://
10309 False nan 13200000 [{'id': 35, 'name': 'Comedy'}, {'id': 18, '… nan
12481 False {'id': 263, 'name': 'The Dark Knight Collec… 185000000 [{'id': 18, 'name': 'Drama'}, {'id': 28, 'n… http://
2843 False nan 63000000 [{'id': 18, 'name': 'Drama'}] http://
292 False nan 8000000 [{'id': 53, 'name': 'Thriller'}, {'id': 80, 'name': 'Crime'}] nan
522 False nan 22000000 [{'id': 18, 'name': 'Drama'}, {'id': 36, 'n… http://
23673 False nan 3300000 [{'id': 18, 'name': 'Drama'}] http://
5481 False nan 15000000 [{'id': 14, 'name': 'Fantasy'}, {'id': 12, … http://
2211 False nan 20000000 [{'id': 35, 'name': 'Comedy'}, {'id': 18, 'name': 'Drama'}] nan
1178 False {'id': 230, 'name': 'The Godfather Collecti… 13000000 [{'id': 18, 'name': 'Drama'}, {'id': 80, 'name': 'Crime'}] nan
1152 False nan 3000000 [{'id': 18, 'name': 'Drama'}] nan
351 False nan 55000000 [{'id': 35, 'name': 'Comedy'}, {'id': 18, '… nan
1154 False {'id': 10, 'name': 'Star Wars Collection', … 18000000 [{'id': 12, 'name': 'Adventure'}, {'id': 28… http://
episode
Format Resize Background color Column min/max Save and Close Close
er Systems.py', wdir='D:/
Warning: Function
moved in 1.2. Please use
LSP Python: ready conda: base (Python 3.9.13) Line 85, Col 51 ASCII CRLF RW Mem 61%